A Multiple Expert-Based Melanoma Recognition System for Dermoscopic Images of Pigmented Skin Lesions

Md. Mahmudur Rahman, Graduate Member, IEEE, Prabir Bhattacharya, Fellow, IEEE, and Bipin C. Desai

***Abstract*— This paper presents an integrated decision support system for an automated melanoma recognition of dermoscopic images based on multiple expert fusion. In this context, the ultimate aim is to support decision making by predicting image categories (e.g., melanoma, benign and dysplastic nevi) by combining outputs from different classifiers. A fast and automatic segmentation method to detect the lesion from the background healthy skin is proposed and lesion-specific local color and texture-related features are extracted. For the classification, combining experts which are classifiers with different structures, are examined as alternative solution instead of an individual classifier. In this approach, probabilistic outputs of the experts are combined based on the combination rules that are derived by following Bayes' theorem. The category label with the highest confidence score is considered to be the class of a test image. Experimental results on a collection of 358 dermoscopic images demonstrate the effectiveness of the proposed expert fusion-based approach.**

I. INTRODUCTION

Skin cancer in the form of malignant melanoma is one of the most common cancers in human being in the world [1], [2]. Detection of malignant melanoma in its early stages considerably reduces mortality, hence this a crucial issue for dermatologists. Because of the worldwide increase of incidence of malignant melanoma reported in the last few decades, researchers attempt to develop noninvasive tools, such as epiluminescence microscopy (ELM) to improve early diagnosis. Dermoscopy or dermatoscopy is the technical name for ELM or skin surface microscopy, a technique that allows in vivo microscopic examination of skin lesions, and it has already proved to be an effective tool for early detection of skin cancers [2]. In clinical practice, several scoring systems and algorithms such as the ABCD rule (ie, asymmetry, border, color, and differential structures), the seven-point checklist, and the Menzies method have been proposed to improve the diagnostic performance of the less experienced clinicians [3]. However, these techniques require formal training and skills in the image interpretation and are highly dependent on the subjective judgment.

Recently, digital imaging and pattern analysis techniques has been found to produce objective and reliable patterns of dermoscopic images [4], [5], [6]. It shows that computer aided diagnosis can be a very helpful tool, particularly in areas which lack experienced specialists. Digitization of the dermoscopic images after the initial visual assessment permits the storage and often used for the comparison when a lesion is being followed over time. Todate, most of the work in the dermatology area have focused on the problem of the skin cancer detection, in which the likelihood of malignancy is computed based on some feature extraction and classification schemes. A variety of statistical and machine learning approaches to classification of dermoscopic images to melanoma, benign or common and dysplastic nevi are currently available [5], [6]. These approaches mainly show the performance comparison of different classifiers independently and discuss their merits and demerits under certain conditions. For example, in [6], the comparison of discriminatory power of the five classifiers, namely k-nearest neighbors (K-NNs), logistic regression, artificial neural networks (ANNs), decision tress, and support vector machines (SVMs) on the task of classifying pigmented skin lesions is analyzed. However, an integrated decision support system based on the fusion of multiple experts or classifiers has not been explored yet in this domain. It has been realized that multiple expert systems which are defined as the instances of classifiers with distinct natures working on distinct feature spaces, can be more robust and more accurate than a single classifier alone [7].

This paper presents a multiple expert fusion-based decision support system for the dermosopic images. By observing the specific characteristics of the images, a segmentation method is proposed for the automatic lesion detection as an image pre-processing step. Lesion specific local color and texture related features are extracted for the image representation and input to the classification systems. For the multiple expert-based fusion, four popular classifier combination techniques (e.g., product, sum, max, and mean) [7] on three different classifier structures (e.g., SVM, Gaussian maximum likelihood (ML), and K-NN) with different input features are adopted as an alternative solution.

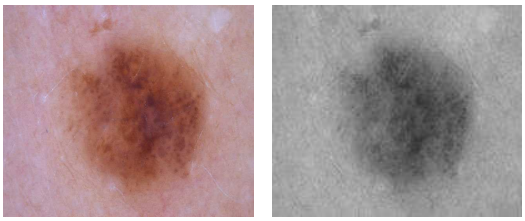
II. SEGMENTATION & LESION DETECTION

Detection of the lesion is a difficult problem in dermoscopic images as the transition between the lesion and the surrounding skin is smooth [4]. Various image segmentation methods have been proposed to delineate the lesion boundaries [4], [8], [9]. For example, an elaborate method is proposed in [8] based on the assumption that color changes from the background to a lesion or from a lesion to the background is more important then the color variations within a lesion or in the background. However,

Manuscript received July 5, 2008. This work was supported by a Postdoctoral Fellowship Fund of Concordia University, Canada

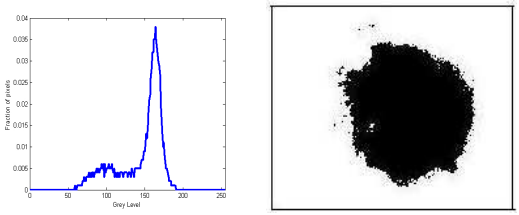
Md. Mahmudur Rahman and Bipin C. Desai are with Dept. of Computer Science and Software Engineering, Concordia University, Montreal, Canada. mah_rahm@cse.concordia.ca

Prabir Bhattacharya is with the Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada prabir@ciise.concordia.ca



(a) (b)

Fig. 1. (a) Original color image (b) Grey level image of the original



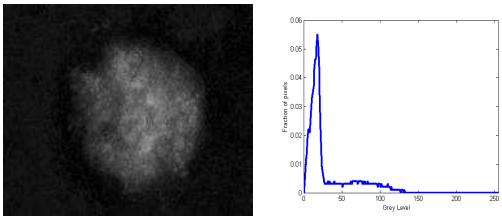
(a) (b)

Fig. 2. (a) Grey level histogram (b) Iterative thresholded image

TABLE I

CORRESPONDENSE BETWEEN HUMAN COLOR PERCEPTION AND NBS DISTANCE

NBS Value	Human Perception
$0 \sim 1.5$	Almost the same
$1.5 \sim 3.0$	Slightly different
$3.0 \sim 6.0$	Remarkably different
$6.0 \sim 12.0$	Very different
$12.0 \sim$	Different color



(a) (b)

Fig. 3. (a) Intensity image generated from HVC space (b) Histogram of the intensity image

the method in [8] has three parameters, which requires user interaction to tune as well as longer processing time. Six different color image segmentation techniques for skin cancer images were compared in [9]. It is found that lowest average error could be achieved by adaptive thresholding and when two or more techniques are combined, the accuracy can be improved further. In accordance with the above observations, we propose a segmentation method by first transforming an image in *RGB* color space to two different intensity images and later combine them to detect the lesion as described in the following sections.

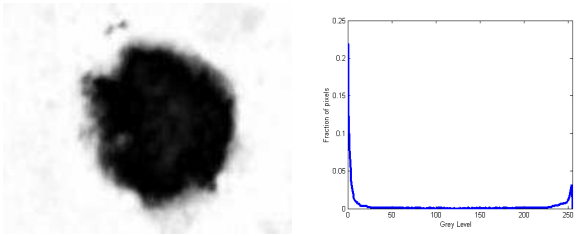
A. Intensity Image Generation from HVC Color Space

The *HVC* color space represents a color in terms of hue (H), which indicates the types of the color, value (V), which tells the total amount of light, and chroma (C) that describes how much white light is mixed with the color (purity). It is famous for its successful imitation of human color perception. There are several ways to mathematically transform the *RGB* to the *HVC* color space. Because *CIE* $L^*a^*b^*$ color space is known for its good perceptual correspondence and simple computation, *RGB* values are first transformed into *CIE XYZ*, and then changed to $CIEL^*a^*b^*$, and then altered to *HVC* values using the formulae in [10].

If $\Delta H, \Delta V$ and ΔC are the differences of H, V, C color components of an image pixel $A = (H_1, V_1, C_1)$ and its background $B = (H_2, V_2, C_2)$, then *NBS* (National Bureau of Standards) color distance between A and B is defined as $Dis(A, B) = 1.2 * \sqrt{2C_1C_2\{1 - cos(\frac{2\pi}{100}\Delta H)\} + (\Delta C)^2 + (4\Delta V)^2}$. There is a close relation between the human color perception and the NBS color distance, which is shown in Table I. Taking advantage of the above properties of the *HVC* space and the *NBS* distance, we transform the image from original *RGB* space to *HVC* space and determine the mean *HVC* values of the pixels from the border (2 pixels wide from each side) of an image. Next, a color image in *HVC* space is transformed into an intensity image in such a way that the intensity at a pixel $f(x, y), 1 \leq x \leq M, 1 \leq y \leq N$ for an image of size $M \times N$ shows the *NBS* color distance of that pixel with the color of the background (e.g., mean of the border). Hence, we obtain an intensity image in which it has higher grey level values in the lesions and lower values in the background after re-scaling as shown in Figure 3(a). We can observe the differences between the grey level histogram in Figure 2(a) of an original image in Figure 1(a) and that of the intensity image in Figure 3(b) generated by the above approach. The later has a more clear separation between the background and foreground pixel intensities with a threshold value of around 27.

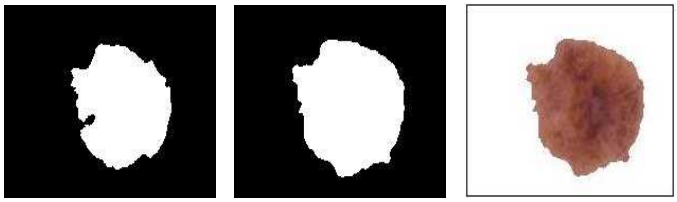
B. Intensity image from fuzzy c-means (FCM) clustering

We also consider another approach to generate an intensity image by utilizing the FCM clustering [11]. FCM is the most widely used fuzzy clustering algorithm which assigns degrees of membership in several clusters to each input pattern. This algorithm is based on an iterative optimization of a fuzzy objective function. The FCM is utilized with the number of clusters is set to two to mainly generate an intensity image for latter processing. In this approach, at each pixel, two membership values are determined where one representing the degree of certainty of a pixel belonging to background normal skin and the other representing the degree of certainty of a pixel belonging to foreground lesion. For input to the clustering, the mean and standard deviation of RGB values are computed as a 6-dimensional feature vector $\mathbf{x}$ in a neighborhood of 5 by 5 pixels around each pixel. After generating the cluster membership values of each pixel, the background membership values are only considered that produces the intensity image after re-scaling



(a) (b)

Fig. 4. (a) Intensity image generated from FCM (b) Histogram of the intensity image



(a) (b) (c)

Fig. 5. (a) Thresholded image from the intensity image of Figure 3(a), (b) Thresholded image from intensity image of Figure 4(a), (c) Final segmented image with lesion mask.

as shown in Figure 4(a). Here also we can observe the differences between the grey level histograms in Figure 2(a) and that of the intensity image in Figure 4(b) with a much deeper and wider valley.

C. Iterative thresholding & post-processing

Thresholding is a computationally inexpensive and fast technique for image segmentation. However, the correct threshold detection is crucial for successful segmentation. We utilize an iterative thresholding approach [12] that basically uses an iterative clustering approach. An initial estimate of the threshold is made (e.g. mean image intensity). Pixels above and below the threshold are assigned to the foreground and background classes respectively. The threshold value is iteratively re-estimated as the mean of the two class means [12]. After computing the threshold values for the two intensity images generated in the pre-processing stage, we obtain the binary images as shown in Figures 5(a) and 5(b). It clearly shows the better separation of the lesion from the background healthy skin in both cases as compared to the image in Figure 2(b). In some cases, the segmentation produces several skin lesion candidates due to the presence of small non-lesion objects. So, a post-processing operation is applied on the binary segmented images to reduce the number of objects based on the morphological operation of opening and closing. Both the opening and closing operations act as nonlinear filters that smooth the contours or lesion border of the input image and tiny artifacts or holes are removed from background and lesion images. After detecting the lesion masks from the segmented images, a simple union (OR) operation is applied to obtain the final lesion mask as shown in Figure 5(c). Usually the largest object is the skin lesion and is thus selected for the feature extraction process.

III. Lesion-Specific Feature Extraction

Suitable local color and texture related features are extracted by considering the mean or average color of the lesion in *HVC* color space and variance-covariances of the color channels by estimating the covariance matrix. The mean color feature vector of a lesion is represented as $\mathbf{x}^{\text{mean}} = [\mu_{\text{H}}, \mu_{\text{V}}, \mu_{\text{C}}]^{\text{T}}$, where μ_{H}, μ_{V} and μ_{C} is the average H, V and C values in HVC space. The cross-correlation among color channels due to the off diagonal of the 3×3 covariance matrix $\mathbf{\Sigma}_j$ is estimated as

$$\mathbf{\Sigma}_j = \frac{1}{N_j - 1} \sum_{k=1}^{N_j} (\mathbf{x}_{j_k} - \mathbf{x}^{\text{mean}})(\mathbf{x}_{j_k} - \mathbf{x}^{\text{mean}})^T \quad (1)$$

where, $\mathbf{x}_{j_k}$ is the color vector of a pixel x_k of the lesion of I_j and N_j is the number of pixels of I_j.

In addition, local texture features are extracted from the grey level co-occurrence matrix (GLCM) [13]. Higher order features, such as energy, maximum probability, entropy, contrast and inverse difference moment are measured based on each GLCM to form a five-dimensional feature vector and finally obtained a twenty-dimensional feature vector $\mathbf{x}^{\text{texture}}$ by concatenating the feature vector of each GLCM. We also uniformly quantized the HVC space into 12 bins for hue (each bin consisting of a range of 30°), 3 bins for the value and 3 bins for the chroma, to generate a 108-dimensional color histogram feature vector $\mathbf{x}^{\text{color}}$. Finally, the color histogram and texture moment based feature vectors are normalized to the zero mean and unit variance and combined or concatenated to form a single vector. Since, the dimension of the combined feature vector is large enough (e.g., 108 for color and 20 for texture for a total of 128) compared to the number of training samples used in the experiments, we applied principal component analysis (PCA) to reduce the feature dimension. The dimensionality of the combined vector is reduced to 10 from 128 where the first 10 eigenvalues related to the 10 principal components (PC's) account for 99.9% of the total variances.

IV. Multiple Expert-Based Fusion

The development of a multiple expert or classifier combination-based system has received increasing attention. The combination of ensembles of ANN and SVM has been studied and evaluated on various image classification data sets involving the classification of digits, faces, objects, etc. [7], [15]. It has been realized that such systems can be more robust and more accurate than using a single classifier alone. The feature vectors of the image representations are often in diversified forms and complementary in nature. It is rather unwise to concatenate them together to form a single feature vector for the input to a classifier. Hence, three classifiers, SVM, Gaussian ML, and K-NN with different input features are considered for expert combination strategies. These classifiers are selected due to their distinct natures of modeling a distribution. Since the outputs of the classifiers are to be used in combination, the modifications are achieved on them to obtain posterior probability values in the range of $[0, 1]$.

A. Classifiers

1) SVM: The SVM [16] classifier is one of the most popular machine learning technologies that has already been used for the melanoma detection [6]. Briefly, we can say that SVM constructs a decision surface between samples of the two classes, maximizing the margin between them. A number of methods have been proposed for the extension of the SVM to multi-class problems. We utilize a multi-class classification method by combining all pairwise comparisons of binary SVM classifiers, known as *one-against-one* or pairwise coupling (PWC) [17]. During the testing of a feature vector $\mathbf{x}$, each classifier votes for one class. The winning class is the one with the largest number of accumulated votes.

For the SVMs training, M number of categories is defined where each category assigns a probability or confidence score to each image as

$$p_m(\mathbf{x}_j) = P(y = m \mid \mathbf{x}_j), \text{for } 1 \leq \mathrm{m} \leq \mathrm{M} \tag{2}$$

The pairwise class probabilities r_{uv} are estimated as an approximation of the original pairwise class probabilities μ_{uv} following the setting of the PWC ensemble in [17]:

$$r_{uv} \approx P(y = u \mid y = u \text{ or } v, \mathbf{x}_j) \approx \frac{1}{1 + e^{A\hat{d}+B}} \tag{3}$$

where A and B are the parameters estimated by minimizing the negative log-likelihood function, and $\hat{d}$ are the decision values of the training data.

2) Gaussian ML: The Gaussian ML classifier is one of the most popular parametric methods of classification in which an image feature with the maximum likelihood is classified into the corresponding class. The likelihood L_m of an image I_j for class label ω_m in terms of input feature vector $\mathbf{x}_j$ is expressed as follows:

$$L_m(\mathbf{x}_j) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_m|^{1/2}} \exp^{[-\frac{1}{2}(\mathbf{x}_j - \boldsymbol{\mu_m})^T \mathbf{\Sigma}_m^{-1}(\mathbf{x}_j - \boldsymbol{\mu_m})]} \tag{4}$$

where, d is the feature vector dimension, $\boldsymbol{\mu_m}$ and $\mathbf{\Sigma}_m$ are the sample mean vector and covariance matrix of the category label ω_m for $1 \leq m \leq M$ and are estimated from a set of training samples. Another way to specify equation (4) is to take the natural logarithms of the quantities involved, such as

$$d_m(\mathbf{x}_j) = -\frac{d}{2}ln(2\pi) - \frac{1}{2}ln(|\mathbf{\Sigma}_m|) - \frac{1}{2}(\mathbf{x}_j - \boldsymbol{\mu_m})^T \mathbf{\Sigma}_m^{-1}(\mathbf{x}_j - \boldsymbol{\mu_m}) \tag{5}$$

where $d_m(\mathbf{x}_j)$ is often called the quadratic discriminant score for the ω_m category. The discriminant scores are mapped to posterior probabilities by the following formula:

$$p_m(\mathbf{x}_j) = P(y = m \mid \mathbf{x}_j) = \frac{\frac{1}{d_m(\mathbf{x}_j)}}{\sum_{j=1}^{M} \frac{1}{d_j(\mathbf{x}_j)}}, \text{for } 1 \leq \mathrm{m} \leq \mathrm{M} \tag{6}$$

3) K-Nearest Neighbor (K-NN): K-NN is another well-known nonparametric statistical classification approach that has been intensively studied for over four decades [14]. Given a feature vector (observation) $\mathbf{x}_j$ belonging to image I_j in the test set, it finds the K nearest neighbors of the observation in the training set. The unclassified observation is then assigned to the class represented by the majority of the K closest neighbors. The outputs of this classifier are converted to the posterior probabilities by applying the following formula:

$$p_m(\mathbf{x}_j) = P(y = m \mid \mathbf{x}_j) = \frac{K_m}{K}, \text{ for } 1 \leq \mathrm{m} \leq \mathrm{M} \tag{7}$$

where K_m denotes the number of nearest neighbors from class label ω_m and K is the total number of nearest neighbors, taken into consideration. K-NN classifiers are especially successful while capturing important boundary details that are too complex for some other classifiers.

For the K-NN classifier, the Bhattacharyya distance measure [14] is utilized in the average color and covariance-based feature spaces. Since, the Bhattacharyya distance metric considers the correlation between the color channels through the covariance matrices, it might perform better then the traditional Euclidean based distance metric. For the SVM and the Gaussian ML classifiers, the color and texture based feature in the PCA sub-space is used as the input feature vector due to the flexibility of the parameter estimation for the Gaussian ML and the utilization of the kernel function in the SVMs.

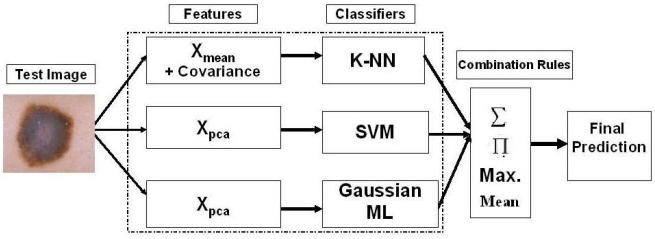


Fig. 6. Process diagram of the multiple expert fusion.

B. Classifier combination

Four popular classifier combination techniques derived from Bayes's theory, such as the product, sum, max, and mean rules [7] are considered for the expert combination strategies. Since the outputs of the classifiers are to be used in combination, the *a posterior* probabilities in the range of $[0, 1]$ for each category will serve this purpose.

In these combination rules, *a priori* probabilities are assumed to be equal and the decision is made by the following formula in terms of the *a posteriori* probabilities yielded by the respective classifiers as

$$\omega_m \Longleftrightarrow \max_{1 \leq m \leq M} p_m^O, \; O \in \{\text{prod}, \text{sum}, \text{max}, \text{mean}\} \tag{8}$$

where the product, sum, max, and mean rules are defined as

$$p_m^{\text{prod}} = \frac{\prod_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j)}{\sum_{m=1}^{M} \prod_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j)} \tag{9}$$

$$p_m^{\text{sum}} = \frac{\sum_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j)}{\sum_{m=1}^{M} \sum_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j)} \tag{10}$$

$$p_m^{\max} = \max_{1 \leq r \leq R} p_r(\omega_m|\mathbf{x}_j) \tag{11}$$

and

$$p_m^{\text{mean}} = \frac{1}{R}\sum_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j) \quad (12)$$

Here, $p_r(\omega_m|\mathbf{x}_j)$ is the *a posteriori* probabilities yielded by an expert r for $1 \leq r \leq R$. In the product rule, it is assumed that the representations used are conditionally statistically independent. In addition to the conditional independence assumption of the product rule, the sum rule assumes that the probability distribution will not deviate significantly from the a priori probabilities. Classifier combination based on these two rules often performs better than the other rules, such as max and mean [7].

The three different classifiers on the different feature spaces are combined or fused by the above rules and finally classify an image to the category with the highest obtained probability value. The process flow diagram of the multiple expert fusion or classifier combination is shown in Figure 6.

V. Experiments & Results

To evaluate the effectiveness of the proposed expert-based decision support system, the experiments are performed on an image database with 358 dermoscopic images collected from two dermatology image atlases [18], [19] with ground truth or known categories. This collection contains pigmented skin lesions of three categories as benign or common nevi (106), dysplastic nevi (118) and melanoma (134). The dysplastic nevi still defines a benign class, but these kind of lesions are the precursors to malignant melanoma, hence classified separately. Majority of the lesions are located always in the central portion of the images to conform with the segmentation algorithm. However, in few cases the segmentation algorithm might fail as the lesions are located all over the images. Since the images are collected from two data sets, they are captured by different devices under different conditions, which makes both the retrieval and the classification tasks even harder. However, in the distributed web-based decision support or teaching systems, this kind of set up is more realistic than the images which are captured from the same hospital or imaging device only.

To experiment with the classification systems, we divide the entire collection of 358 images as the training and test sets. The training set for the classifiers is 40% of the entire collection and remaining 60% is used as the test set. Hence, of the 144 training set images, 42 is made up of benign, 48 dysplastic and 54 melanoma category; the remaining 214 images are the test set. This training set is utilized to perform the PCA of the combined color and texture feature, training and parameters estimation for the SVM, Gaussian ML and K-NN classifiers.

A. Classification Accuracies

To evaluate the multiple expert-based system, at first we need to train the individual classifiers and estimate various parameters. For the SVM-based image classification, recent work shows that the radial basis kernel function (RBF) works well when the relation between class labels and attributes is nonlinear. Therefore, we use the RBF kernel as a reasonable first choice. There are two tunable parameters while using the RBF kernel : C and γ. It is not known beforehand which C and γ are the best for the classification problem at hand and are selected by cross-validation (CV). In the training stage, the goal is to identify the best (C and γ), so that the classifier can accurately predict testing data. For training set, a 10-fold CV is conducted, where we first divide the training set into 10 subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining 9 subsets. Thus, each instance of the whole training set is predicted once so the CV accuracy is the percentage of data which are correctly classified. Hence, we performed the following procedure at the training stage:

- Considered the radial basis kernel function (RBF) $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma|\mathbf{x}_i - \mathbf{x}_j||^2), \gamma > 0$.
- Used 10 fold cross-validation to find the best parameter C and γ by measuring the cross-validation classification accuracy.
- Used the best parameter C and γ to train the whole training set to generate a model file.

After 10-fold CV, a training accuracy (% of images correctly classified) of 70.91% is achieved with the best parameters, C and γ as shown in Table II. In the experiments, we also used polynomial kernels with degree 1 and 2 with a cost factor of $C = 10$. In all the cases, the CV accuracies are presented in Table II. We have utilized the LIBSVM software package [20] for the implementation of the SVM classifiers.

To experiment with the Gaussian ML classifier, for off-line parameter estimation from the training samples, we estimated means (μ) and covariance matrices (Σ) for three different categories in the reduced PCA sub-space using equation (4) from the training image set. For K-NN classifier, we experimented with $K = 20$ closest neighbor by applying the Bhattacharyya distance metric, where 20 closest neighbor is a reasonable one to provide a statistical significance.

TABLE II
Cross Validation Accuracy (SVM)

Kernel	C	γ	Degree	Accuracy
RBF	200	.005		70.91%
Polynomial	10		1	69.36%
Polynomial	10		2	67.15%

TABLE III
Performance Comparison of Single Classifiers

Classifier	Accuracy
SVM (RBF)	72.20%
SVM (Polynomial-1)	70.22%
SVM (Polynomial-2)	70.02%
K-NN	61.87%
Gauss ML	56.59%

TABLE IV

PERFORMANCE COMPARISON OF DIFFERENT EXPERT COMBINATIONS

Prod	Sum	Max	Mean
75.23%	75.69%	74.29%	71.75%

TABLE V

CONFUSION MATRIX FOR EXPERT COMBINATION ON SUM RULE

True/Assigened	1	2	3	Accuracy
1 (benign)	40	11	13	62.50%
2 (dysplastic)	8	54	8	77.14%
3 (malignant)	6	7	67	83.75%
Overall				75.69%

After training and estimating the parameters of the classifiers, we measured the accuracy of individual classifiers with the independent test set. From Table III, it is clear that SVM classifier performed better with both RBF and polynomial kernel compared to other classifiers (e.g., K-NN and Gaussian ML). However, for later classifier combination, only SVM with RBF kernel is utilized with other classifiers as it performed better in the test set as shown in Table III. The accuracy of the Gaussian ML classifier is not satisfactory, which might be due to the inaccuracy in parameter estimation from the small sample size. When we have applied the classifier combination rules involving all the three classifiers (e.g., SVM, K-NN and Gaussian ML), the improved accuracies are observed in three out of the four cases as shown in Table IV. Except the mean rule, all other combination rules performed better then the best performed individual classifier (e.g., SVM with RBF kernel) as shown in Table IV. The confusion matrix for the test set is calculated by the best combination rule (e.g., sum rule) and displayed in Table V. From this Table, we can observe that the good classification accuracy (83.75%) is obtained by the malignant category. This is important from the diagnostic viewpoint as false classification of melanoma has deadly consequences for the patients. Overall, we achieved an accuracy improvement of more than 3% based on the multiple expert-based fusion as compared to the best performance of the single classifier (e.g., the SVM with an accuracy of 72.20%).

VI. CONCLUSION

We have presented a multiple expert-based system as a diagnostic aid for melanoma recognition. The experimental results indicate that the system is effective to predict the categories of images for diagnostic correctness. We plan to incorporate more advanced features related to the diagnostic relevance into our system and experiment with other classification and combination techniques as well. However, the presence of an expert dermatologist is considered necessary for the overall visual assessment of the skin lesion and the final diagnosis based on objective evaluation suggested by the system and contextual information from the patient data and the histopathological tests.

VII. ACKNOWLEDGMENT

The authors would like to thank C. C. Chang and C. J. Lin for the LIBSVM software tool [20] that is used for the SVM-based classification purpose.

REFERENCES

[1] D.S. Rigel, and J.A. Carucci, "Malignant melanoma: prevention, early detection, and treatment in the 21st century," *CA: Cancer J Clin.*, vol. 50, pp. 215-36, 2000.

[2] M. Binder, M. Schwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, and H. Pehamberger, "Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists," *Arch Dermatol.*, vol. 131(3), pp. 286–291, 1995.

[3] R.H. Johr, "Dermoscopy: alternative melanocytic algorithms-the ABCD rule of dermatoscopy, Menzies scoring method, and 7-point checklist," *Clin Dermatol.*, vol. 20(3), pp. 240-247, 2002.

[4] H. Ganster, A. Pinz, R. Röhrer, E. Wildling, M. Binder, and H. Kittler, "Automated Melanoma Recognition," *IEEE Trans Med Imaging*, vol. 20, No. 3, pp. 233-239, 2001.

[5] M. Binder, H. Kittler, A. Seeber, and A. Steiner, "Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network," *Melanoma Res.*, vol. 8, pp. 261-266, 1998.

[6] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder, "A comparison of machine learning methods for the diagnosis of pigmented skin lesions," *J Biomed Inform.*, vol. 34(1), pp. 28–36, 2001.

[7] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans Pattern Anal Machine Intell*, vol. 20 (3), pp. 226–239, 1998.

[8] L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines, C. Yu, A. Dhawan, and A. Huntley, "Segmentation of skin cancer images," *Image Vis. Computing*, vol. 17, pp. 65-74, 1999.

[9] S.E. Umbaugh, R.H. Moss, W.V. Stoecker, and G. A. Hance, "Automatic color segmentation algorithms: With application to skin tumor feature identification," *IEEE Eng. Med. Biol. Mag.*, vol. 12(3), pp. 75-82, 1993.

[10] M. Miyahara, and Y. Yoshida, "Mathematical transform of (r,g,b) color data to Munsell (h,v,c) color data," *SPIE Proc. in Vis Comm and Image Process.*, vol. 1001, pp. 650–657, 1988.

[11] J.C. Bezdek, M. R. Pal, J. Keller, and R. Krisnapuram, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing.*, Kluwer Academic Publishers, Boston, 1999.

[12] T.W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-8, pp. 630-632, 1978.

[13] R.M. Haralick, Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans Syst Man Cybernetics*, vol. 3, pp. 610-621, 1973.

[14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Boston, 2nd edition: Academic Press; 1990.

[15] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. Sys. Man. Cybern.*, vol. 23(3), pp. 418-435, 1992.

[16] J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, 1998.

[17] T.F. Wu, C.J. Lin, and R.C. Weng, "Probability Estimates for Multi-class Classification by Pairwise Coupling," *J Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[18] Dermatology Image Atlas, "Dermnet: the dermtologist's image resource", 2006, available at: http://www.dermnet.com/

[19] S.W. Menzies, K. Crotty, C. Ingvar, and W. McCarthy, *An atlas of surface microscopy of pigmented skin lesions dermoscopy*, 2nd Edition & CDROM quiz, Sydney, McGraw-Hill Book Co., 2003.

[20] C.C. Chang, and C.J. Lin, "LIBSVM: a library for support vector machines.", 2001, Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm